AD6[illegible]0239

AN EMPIRICAL TEST OF EXPONENTIAL SMOOTHING

Max Astrachan and Craig C. Sherbrooke

MEMORANDUM
RM-3938-PR
MARCH 1964

AN EMPIRICAL TEST OF EXPONENTIAL SMOOTHING

Max Astrachan and Craig C. Sherbrooke

This research is sponsored by the United States Air Force under Project RAND—contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA

PREFACE

This Memorandum is another in the series of RAND studies concerned with predicting demands for spare parts. Our objective has been to determine empirically whether exponential smoothing techniques can predict demands more accurately than the moving average techniques currently being used in the Air Force. We applied various forms of these two types of prediction procedures to three sets of historical data:

1) Base demands for Hi-Valu and Category II-R items for the B-52,

2) Components of the Falcon missile, and

3) Depot issues for low cost Category III items on the B-52.

The Navy and some industrial firms are already using exponential smoothing techniques. We have undertaken this study to examine the possible usefulness of these techniques to the Air Force. Personnel who are concerned with predicting the demand for spare parts should be especially interested in exponential smoothing methods.

SUMMARY

This Memorandum, a comparative study of techniques for predicting the demand for spare parts, attempted to discover the potential advantages which exponential smoothing has over the moving average procedures the Air Force is now using.

In exponential smoothing the predicted average is found by weighting the average computed at the end of the last time period with the observed demand during the current period. One may vary the weighting constant on the basis of how much weight one wants to put on the last average.

Various forms of exponential smoothing and averaging were applied to three sample sets of data: Hi-Valu and Category II recoverable B-52 parts, components of the Falcon missile, and Category III depot issues of B-52 items. To the usual methods for selecting preferred techniques used in previous studies we added a loss function, an aggregate measure of accuracy which balances procurement costs against holding costs.

The study led to the following findings and conclusions:

(1) For any of the three sets of data, exponential smoothing was not a significantly better prediction technique than the cumulative issue rate procedures now being used in the Air Force. Nevertheless, it does have definite computational advantages which may be valuable. In first order smoothing only one average need be stored for each item. The rate of response due to the smoothing constant can be easily changed, and trends can also be accommodated readily.

(2) A measure of aggregate loss, such as the loss function introduced in this study, should be used to select preferred smoothing

techniques. The ranking procedures used in earlier studies (as well as in this one) do not always serve to discriminate among techniques. More important, they ignore the magnitude of the errors.

(3) The use of program element information for the Falcon components improved the accuracy of our predictions; application of requisition data, which was available for the Category III items, did not.

(4) With any of the techniques applied to the Category III items, prediction accuracy did not increase substantially when the base period was made longer than one year.

ACKNOWLEDGMENTS

We are indebted to Warren G. Briggs for a number of suggestions in connection with the first part of our study, and to George J. Feeney for the measure of aggregate error described in Section IX.

CONTENTS

I. INTRODUCTION

PURPOSE AND SCOPE

This Memorandum is one of a continuing series on the prediction of demand for spare parts, a subject which is a perennial one in Logistics. Our scope here does not involve testing a complete inventory policy. Rather, we are restricting ourselves to the simpler question of prediction accuracy.

We have chosen such a limited objective for two reasons. In the first place, exponential smoothing is a special case of a general class of prediction techniques the peculiar properties of which make it a convenient technique for application. Secondly, it is currently being used by the Navy and some industrial firms, and we feel that the Air Force should be interested in the evaluation of a prediction technique that other users have found simple and successful.

OUTLINE OF THE STUDY

In Section II we present a general discussion of the problem of predicting spares demand. Section III describes the moving average and exponential smoothing techniques. Two fundamentally different sets of data were employed. The first set of data, from two bases on 272 B-52 items and 27 Falcon recoverable parts, is introduced in Section IV; the test design for this data is found in Section V. Section VI deals with measures of accuracy, and Section VII presents results. In Section VIII, the second set of data, covering ten quarters of depot issues for 425 Category III items, is described and results are given. Section IX develops a method of measuring error for an aggregation of items; this method is then used to evaluate prediction

techniques for the Category III items. Findings and conclusions constitute Section X.

II. PROBLEM OF PREDICTING SPARES DEMAND

ASSUMPTIONS

We do not pretend to offer an exhaustive discussion of the problem of predicting spares demand. Instead, making several assumptions about the real world, we shall define and turn our attention to one specific problem. We will define demand and specify the form of the demand data. Then we shall discuss some of the relevant considerations in demand prediction that influence our research design.*

Definition of Demand

We are interested only in "recurring" demand for spare parts, i.e., all except one-time demands, such as technical order compliance demands where a modified part is substituted for an inferior but serviceable one. In the recoverable parts area, the definition of demand is further restricted to exclude a demand made on the supply system if the bench check shows that the reparable turn-in is serviceable. The definition of a demand for a recoverable spare part is thus identical with the Air Force definition of a maintenance replacement removal (MRR).

Demand Data

We shall limit our interest to problems in which the demand data is summary information specifying total demand for each period by line item. In the language of the mathematician we have a time series of

---

*Some of the relevant ideas are also discussed in A Comparative Study of Prediction Techniques, by Max Astrachan, Bernice Brown, and J. W. Houghten, The RAND Corporation, RM-2811, December, 1961.

demand. Initial estimates may or may not be available.* In some cases the demand data may also include a time series of requisition data, e.g. the total number of demands and the total number of requisitions used in making the demands by quarter and item.

Program Element Data

Auxiliary data such as the flying hours per month may be on hand. In Air Force demand prediction problems a variety of program elements has been examined, including sorties, low-level flying hours, and equipment hours.** The objective is always to find some program element or elements that will transform the original time series of data into a new series of data from which more accurate predictions can be made.

Probability Distribution of Demand

We shall not make any assumption about the "true" probability distribution of demand nor shall we restrict ourselves to stationary demand.***

Objective of Demand Prediction

Our objective is to find a technique that will predict demand most accurately over a reasonably long period of time, such as procurement leadtime -- e.g. nine months or a year. These predictions

---

*In the event that initial estimates and demand data are both available, one technique for combining them is suggested in W. H. McGlothlin, Development of Bayesian Parameters for Spare Parts Demand Prediction, The RAND Corporation, RM-3699, July, 1963.

**See H. S. Campbell, The Relationship of Resource Demand to Airbase Operations, The RAND Corporation, RM-3428-PR, January 1963.

***If demand is stationary the minimum variance estimate of mean demand is obtained by weighting all past data equally.

cannot be translated directly into procurement actions because safety levels would have to be added for demand and leadtime variations.

GENERAL CONSIDERATIONS

We have defined the problem above. Let us examine here the rationale for our choice of problem. We assumed a time series of summary data, because this is the sort most commonly encountered. We set an objective of predicting demand over a procurement leadtime so that the prediction errors could be computed. In contrast, a prediction of average demand plus a safety level, though appropriate for procurement, would bias the error distribution. Furthermore, the safety level is an additive to the average demand and can be considered separately.

Once one has selected the objective of predicting demand over a procurement leadtime, it is not necessary to assume a specific probability distribution. In statistical terms, the estimation of the mean value does not require that we know the form of the probability distribution. Furthermore, we have good reasons to avoid the selection of a specific probability distribution. In the first place, we are interested in the prediction of demand at different echelons including base and depot. Variance of demand at each echelon is strongly influenced by the requisitioning policies at lower echelons. For example, if bases order a year's supply of an item instead of a quarter's worth, the variance of demand on the depot will be increased. If one probability distribution were to suffice for all echelons, it would be essential that the distribution have at least two parameters -- to provide for different means and variances by line item. Secondly, we

believe that a demand prediction technique should be sensitive to slow changes in the mean demand rate. In other words, our examination of data leads us to believe that items show non-stationary demand characteristics that should not be neglected. The estimation of parameters for a two-parameter non-stationary probability distribution, however, is a formidable task.

III. PREDICTION TECHNIQUES

BACKGROUND

The current USAF procedure for computing the "historical usage rate" element of the requirements computation at both the base and the depot levels is essentially an unweighted moving average which is updated periodically. Such averages have many of the desirable characteristics of a practical method for smoothing out the fluctuations in a demand history to get an estimate of the expected demand rate. They have a stable response to changes, and the rate of response can be controlled by the number of months (or observations) included in the average. Although moving averages are simple to compute, they require that the individual observations used in computing be retained so that new information can be added and old information dropped. Exponential smoothing or exponential weighting is similar to a moving average, except that all observations are used. The former, however, does not require the keeping of a long historical record in the active file or computer, and the data-processing requirements are therefore decreased. Like the ordinary moving average, exponential smoothing has a stable response to changes, but the rate of response can be readily adjusted. Then, too, the method can be extended to the calculation of trends, and changes in trends, with very little extra data-processing.

MOVING AVERAGES

Before defining exponential smoothing, let us recall the procedure followed in updating a moving average. Suppose we have observed d_t demands in the current time period, d_{t-1} demands in the last period,

d_{t-2} two months ago, etc. Then the updated average demand at the end of the current period is given by

$$D_t = \frac{1}{N}(d_t + d_{t-1} + d_{t-2} + \cdots + d_{t-(N-1)})$$

$$= D_{t-1} + \frac{1}{N}(d_t - d_{t-N}) ,$$

where D_{t-1} is the value of the moving average at the end of the last period. Thus the updating is accomplished by adding to the prior average a fraction of the difference between the current observation and the observation N periods old, d_{t-N}. This is the effect of "adding the newest and discarding the oldest, then averaging the result."

In this technique the most recent and the oldest observations have the same influence (weight) on the updated average. In fact, each of the observations in the N periods has the same weight, 1/N. The moving average obviously requires that the individual observations for all N periods be retained and used for each updating.

A method that could be used to avoid the equal weighting of all data regardless of age is to use a weighted moving average. A sequence of positive weights, a_0, a_1, a_2, ... a_{N-1}, whose sum is one, is arbitrarily selected. The updated average is then

$$D_t = a_0 d_t + a_1 d_{t-1} + a_2 d_{t-2} + \cdots + a_{N-1} d_{t-(N-1)} .$$

The weights can be selected so as to give more consideration to current than to earlier data. This method, however, involves substantially more computation than an unweighted moving average, and also requires that all N observations be retained. In addition, although the weights

are arbitrary, subject only to the condition that they are positive and their sum is unity, N must be preassigned.

A linear trend or higher order model can be accommodated by a least squares procedure in which each squared error is weighted by the appropriate a_i.

EXPONENTIAL WEIGHTING

The Basic Concept. Suppose that we had stored only the average demand computed last month* and had not stored the individual observations. This month we have a new demand quantity and want to update the average. It seems logical that if the demand this month is higher than the stored average, we should increase the latter. Conversely, if the number of demands observed this month is smaller than the average at the end of the previous month we should decrease it. Furthermore, if the difference is small, the adjustment should be small. If the demand has been substantially above (or below) the stored average, the new estimate should be increased (or decreased) by a sizeable amount.

The exponentially-weighted average computation can be described symbolically as follows:

$$(1) \qquad D_t = D_{t-1} + \alpha(d_t - D_{t-1}), \qquad 0 < \alpha < 1$$

or by rearrangement of terms,

$$(2) \qquad D_t = (1-\alpha)\, D_{t-1} + \alpha d_t,$$

*Although we are describing exponential weighting in terms of demands per month, the procedure can be used to update any type of data sequence or discrete time series such as demands per 1000 flying hours, failures per 100 checkouts, issues, etc. The time period could also be a day, month, quarter, etc.

where:

D_t = updated average (prediction) made at end of current period, t;

D_{t-1} = updated average at end of prior period, t-1;

d_t = observation (i.e., demand, demand rate, etc.) for the current period; and

α = weighting or smoothing constant, a value between zero and one.

Repeated application of (2) to the most recent N observations gives

$$(3)\qquad \begin{aligned} D_t &= \alpha d_t + \alpha(1-\alpha)d_{t-1} + \alpha(1-\alpha)^2 d_{t-2} + \cdots \\ &\quad + \alpha(1-\alpha)^{N-1} d_{t-(N-1)} + (1-\alpha)^N D_{t-N} \\ &= \alpha \sum_{n=o}^{N-1} (1-\alpha)^n d_{t-n} + (1-\alpha)^N D_{t-N}, \end{aligned}$$

where D_{t-N} is the prediction at the end of period t-N, or at the start of these N observations. This value could also be considered as the initial estimate for D prior to any experience.

The weight assigned to each observation is a constant α times a fraction $1-\alpha$ with exponent equal to the age of that observation -- hence the terms "exponential weighting" or "exponential smoothing." As N becomes very large, i.e., as we have a very large number of observations on which to base D_t, the "initial estimate" term drops out (i.e., "adequate" actual experience becomes available). The sum of the exponential (literally "geometric" in this discrete case) weight approaches one:

(4) $$D_t = \lim_{N \to \infty} \left[\alpha \sum_{n=0}^{N-1} (1-\alpha)^n d_{t-n} \right] + 0 ,$$

(5) $$\text{Sum of weights} = \alpha \sum_{n=0}^{\infty} (1-\alpha)^n = \alpha \left[\frac{1}{1-(1-\alpha)} \right] = 1.$$

Figure 1 shows graphically the weight assigned to data t periods old for three values of α. It can also be seen from Eq. 3 that the total weight given to all observations prior to the N most recent ones is $(1-\alpha)^N$.

It is obvious from Eqs. 1 through 3 above that the "responsiveness" or "sensitivity" of the prediction (D_t) to current data (d_t) depends upon the magnitude of the constant, α. Larger values of α give additional weight to the more recent observations; the converse is true for smaller values. Yet, all data are always considered in D_t, however trivial the weight may be. This aspect of the exponential weighting procedure actually leads to drastic simplicity in data storage requirements. As can be seen from Eqs. 1 and 2, the only historical information needed at each updating is the prior period's prediction, D_{t-1}.

This data storage requirement contrasts sharply with the N period (N is often 12 or 24 months) unweighted moving average. At each updating in this procedure the oldest observation is discarded and the most recent one added to compute the new average. This means, of course, that the values for each of N observations must be stored.

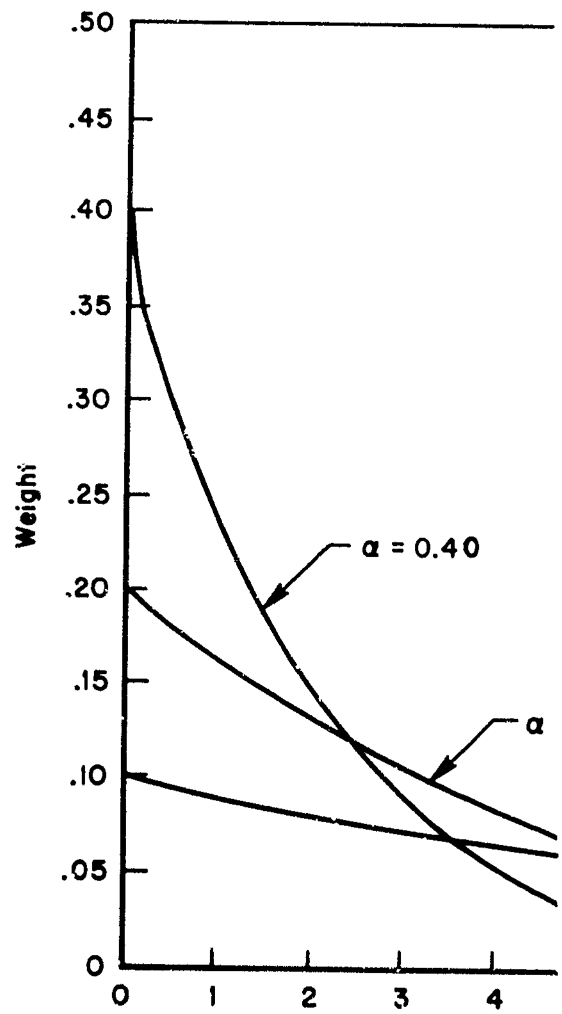
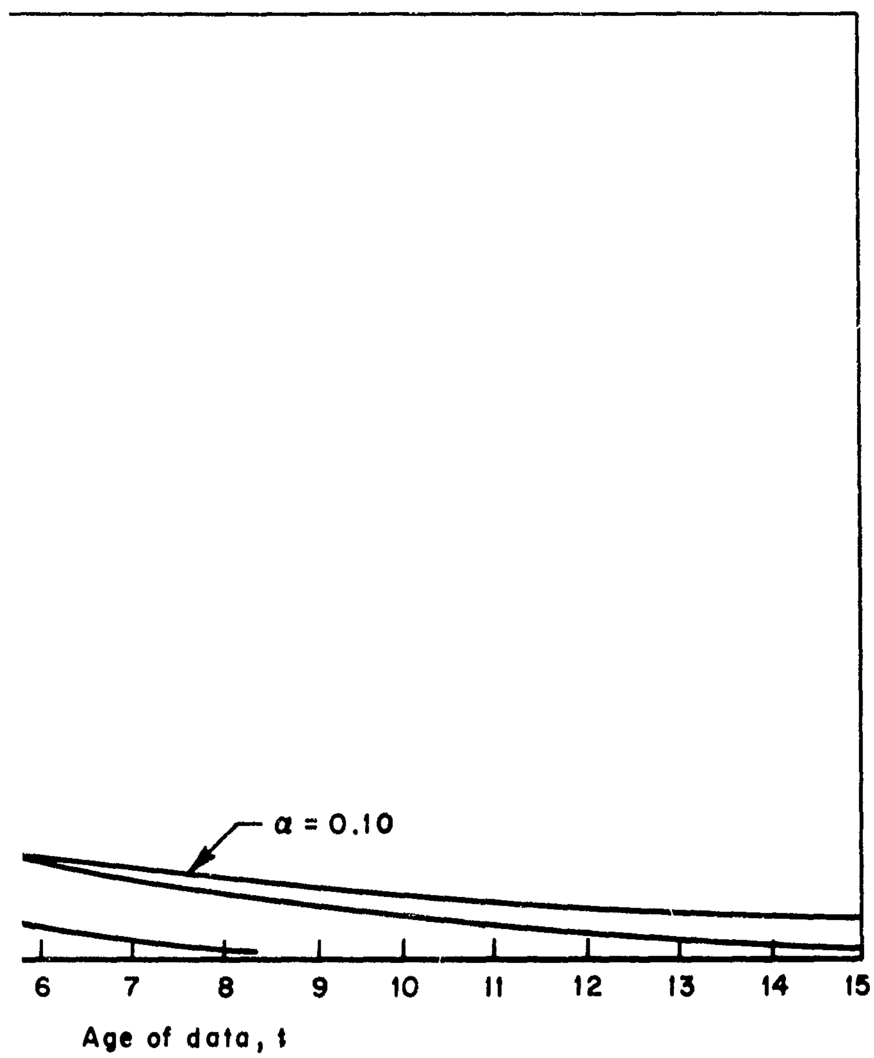


Fig. 1 — Weight assigned to data



iods old for three values of the weighting constant, α
veight $= \alpha(1 - \alpha)^t$)

BLANK PAGE

<u>Selection of α</u>. The rationale behind selecting an appropriate value for α is no more complex, and unfortunately no easier to justify, than that behind selecting N for the commonly used moving average. Either case requires a compromise between 1) promptly reflecting true changes in the data sequence, and 2) avoiding excessive response to mere chance fluctuations. A large value for α improves the rate of response to a changing pattern in the data sequence by giving more weight to recent data. The same is true, of course, with a small value of N. In both situations, however, the ability of the technique to smooth out random fluctuations is decreased.

The true optimum value for N or α can never really be determined for such a complex prediction problem as encountered in large scale logistics systems. An extensive test against past experience, like that reported herein, can indicate which of several trial values of α would have been best for the data studied. But one cannot be certain that future data would lead to the same conclusion.

One approach for establishing test values of the smoothing constant is through an equivalence to traditionally accepted N-period unweighted moving averages. The data included in the latter have an average age equal to:

$$\frac{0 + 1 + 2 + \ldots + (N-1)}{N} = \frac{N-1}{2} . \tag{6}$$

Using the weights shown in Eq. 4, the data in the exponentially-weighted average have an average age equal to:

$$(7) \qquad \alpha(0) + \alpha(1-\alpha)(1) + \alpha(1-\alpha)^2(2) + \ldots + \alpha(1-\alpha)^n(n) + \ldots$$

$$= \alpha \sum_{n=0}^{\infty} n(1-\alpha)^n = \frac{1-\alpha}{\alpha} .$$

If we now define an exponential weighting procedure as being equivalent to an N-period unweighted moving average if the data have the same average age, then by equating Eq. 6 with Eq. 7, we get

$$(8) \qquad \alpha = \frac{2}{N+1} , \quad \text{and } N = \frac{2-\alpha}{\alpha} .$$

Another means of defining equivalence is to equate the variances of predictions made by each technique. If σ^2 is the variance of the observed data, then the variance of the N-period moving average is σ^2/N, assuming the observations to be independent and from the same population. The predictions generated by exponential smoothing can be shown* to have variance equal to $\alpha\sigma^2/(2-\alpha)$. Equating the two variances gives the same results as in Eq. 8 above.

Table 1 shows some of the equivalent values of N and α. We see that a value of α equal to 0.200 gives results equivalent to those obtained from an unweighted moving average of 9 periods in the sense described above. In the latter, the 9 most recent observations are used, and the older ones are discarded. In exponential weighting, all the observations are used. As can be seen from Eq. 3, the total weight assigned to all observations older than the most recent nine is $(1-0.20)^N = (.80)^9 = .107$. Because of this difference between the two techniques, predictions may differ substanticlly although the

---

*See, for example, R. G. Brown, Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice-Hall, Englewood Cliffs, N.J., 1963, p. [illegible].

techniques are equivalent in terms of the average age of the data and the prediction variance.

Table 1

VALUES OF THE WEIGHTING CONSTANT CORRESPONDING TO AN EQUIVALENT MOVING AVERAGE

Number of Periods in Moving Average (N)	Equivalent Exponential Weighting Constant (α)
3.0	0.500
6.0	0.286
9.0	0.200
12.3	0.150
19.0	0.100
24.0	0.080

<u>Linear Trend and Higher Order Models</u>. Up to this point we have been discussing only the simplest prediction model, namely, the straightforward extrapolation of an updated constant value. Sometimes a data sequence is more realistically predicted by a more elaborate model, such as one including a linear or quadratic trend or a cycle. We shall be concerned here with only the first of these.

If there is no trend in the data so that a simple average a_t can be used for extrapolation, then, as we have seen, single exponential smoothing gives the estimate $a_t = D_t$. If there is a linear trend in the series, the model is assumed to be of the form

(9) $$\hat{d}_{t+k} = a_t + kb_t,$$

where

$\hat{d}_{t+k}$ = prediction for period t+k, k=1, 2, ... ,

b_t = trend rate as of the end of period t, expressed in units of increase or decrease per period,

a_t = trend value as of the end of period t.

The prediction of total demand for the next m periods is

$$\sum_{k=1}^{m} \hat{d}_{t+k} = ma_t + \frac{m(m+1)}{2} b_t .$$

The model in Eq. 9 requires updating two constants, a_t and b_t. This can be done by use of "second-order smoothing". A second-order smoothed (or exponentially-weighted) average is simply an average of the averages, so to speak. The computation is identical to that in Eqs. 1 through 4 for first order smoothing, only in this case the data sequence is the series of D_t rather than d_t values. Thus if D_t' denotes the second-order exponentially weighted average up-dated at period t, then

$$(10) \qquad D_t' = (1-\alpha)D_{t-1}' + \alpha D_t$$
$$= \lim_{N\to\infty}\left[\alpha \sum_{n=0}^{N-1} (1-\alpha)^n D_{t-n}\right] = \alpha \sum_{n=0}^{\infty} (1-\alpha)^n D_{t-n} ,$$

where D_t is the first-order average from Eq. 1.

Estimates of a_t and b_t are given by[*]

$$(11) \qquad \hat{a}_t = 2D_t - D_t', \quad \text{or} \quad D_t + \frac{1-\alpha}{\alpha} \hat{b}_t$$

$$\hat{b}_t = \frac{\alpha}{1-\alpha}(D_t - D_t'), \quad \text{or} \quad D_t' - D_{t-1}' .$$

---

[*] R. G. Brown and R. F. Meyer, "Fundamental Theorem of Exponential Smoothing", Journal of the Operations Research Society of America, Vol. 9, No. 5, Sept.-Oct., 1961, pp. 673-687.

Note that both $\hat{a}_t$ and $\hat{b}_t$ can be written in terms of the first and second-order smoothed values of the observed data sequence. The latter then are the only values which need be stored.

Estimates of a_t and b_t can be derived informally if one assumes that demand follows a linear model exactly, plotting the response of D_t and D'_t and choosing $\hat{a}_t$ and $\hat{b}_t$ so that the prediction coincides with the assumed linear demand. A rigorous proof* has been given to show that the coefficients of any polynomial model of degree k can be expressed in terms of the first k + 1 degrees of exponential smoothing, and that this polynomial minimizes the exponentially weighted least square error.

In other words the coefficients of the polynomial model can be estimated by a least squares procedure in which the squared errors are exponentially-weighted. This is completely analogous to the procedure discussed above for moving averages, but here only k + 1 smoothed averages rather than all N observations need be retained.

---

*Brown and Meyer, op. cit. See also J. M. Dobbie, "A Simple Proof of a Theorem in Exponential Smoothing," Journal of the Operations Research Society of America, Vol. 11, No. 3, May-June, 1963, pp. 461-463.

IV. THE DATA - DESIGN 1

This test of exponentially-weighted techniques is essentially a sequel to another study[*] of demand prediction techniques in the sense that we are using some of the same data and procedures. Two different test designs are employed in the present study. They involve different sets of data and techniques to be compared. Those used in Design 1 are described in this and in Section V. Design 2 is described in Section VIII.

The data used in Design 1 are from the above mentioned study, which focused upon a sample of B-52 parts and Falcon components. Since a complete description of these data is given there, we shall include only a brief review here.

The sample of B-52 data consists of Hi-Valu and Category II recoverable parts from six major property classes: Engine Components, Airframe Structural Components, Gunnery Components, Bombing Fire Control Components, Communications Equipment, and Aircraft Accessories. The data, from two bases, Loring and Castle, cover a period of 33 months, from January, 1956 through September, 1958. RM-2811 (pp. 33-40) describes how a sample of 272 parts was selected from the original group of 7500 part numbers. This sample contains only parts with 5 or more demands during the 33-month experience period.

These spare parts consumption data are related to a total of 68,000 flying hours of operation of 169 different B-52's. These planes include series B, C, D, E, F. Not all spare parts were applicable to all series. The earlier study took this into account along

---

[*]Astrachan, Brown, and Houghton, *op. cit.*

with the flying hour program of each aircraft series. We decided that the 175 (of the 272) parts which are applicable to all series would be adequate for this test. A further restriction was placed on the data for this study by our using only those line items (125 of the 175) which had some demand during the first 21 months. This was the maximum experience period used for prediction, and it seemed pointless to use any of these prediction techniques on the remaining 50, since these techniques always predict zero when based on zero demands. The final B-52 sample for this study, then, consisted of 125 parts which were applicable to all series and which had some demand during the first 21 months.

Data for the 27 Falcon components* cover a period of 26 months, from May, 1955 through June, 1957. About 30,000 missile checkouts (the program element used for prediction) were performed during this period. We used the 23 (of the 27) components which had some demand during the first 14 months -- the maximum experience period used for prediction in this case. RM-2811 describes the Falcon data in greater detail than we do here.

---

*These are all the components of the missile.

V. THE TEN PREDICTION TECHNIQUES - DESIGN 1

This section outlines the specific tested applications of the general concepts described earlier. The ten prediction techniques* used are designed to compare the following: 1) exponential smoothing with and without a program element, with and without trend; 2) the effect of different values of the weighting constant α; 3) cumulative average; 4) nine-month moving average; and 5) where appropriate, these procedures as against those in RM-2011.

A preliminary trial run of the exponentially-weighted average technique was used to examine a wide range of values for α. The results indicated that for our data, values of 0.10 and 0.20 would be most appropriate in the sense of yielding reasonably accurate, yet significantly different, results. This finding is consistent with the values for α currently being used by the Navy and several industrial firms.

Techniques 1-4 are exponentially-weighted techniques using as inputs demands per program element -- flying hours for the B-52 parts, checkouts for the Falcon components. For Technique 1 we used $\alpha = 0.10$. The value of the updated demand rate D_t for a given part was computed using t months of experience and Eq. 2. Then, assuming that this part would continue to be demanded at this same rate in the future, we obtained the forecast for a particular month by multiplying D_t by the actual activity (flying hours or checkouts) for that month.** For

*Although we refer to these as different techniques, they are really variations of two basic models -- exponential smoothing and averaging.

**In a realistic prediction problem, the actual flying hours for future months would have to be estimated.

each experience period, predictions were made for the following 12 months within the limits of the data. Thus, there are 21 such prediction sets (each consisting of 12 monthly predictions) for the B-52 parts and 14 sets for the Falcon components.

Technique 2 is the same as Technique 1 except that we used $\alpha = 0.20$.

In Technique 3, we assume that there was a linear trend in the demands per program element during the first t months of experience and that this same trend will continue in the future. Using $\alpha = 0.10$, we compute $\hat{a}_t$ and $\hat{b}_t$ from Eq. 11, and $\hat{d}_{t+k}$ from Eq. 9 for k = 1, 2, ... 12. The forecast for a particular month is obtained by multiplying the appropriate $\hat{d}_{t+k}$ by the activity (flying hours or checkouts) for that month. Again there are 21 predictions sets for the B-52 parts and 14 sets for the Falcon components.

Technique 4 is the same as Technique 3 except that we used $\alpha = 0.20$. Note that second order smoothing must be used in Techniques 3 and 4.

Technique 5 is a simple unweighted cumulative average or issue-rate technique (identical to Technical I of RM-2811). In this technique, the total number of demands for a given part during t months of experience is divided by the total activity (flying hours or checkouts) during that period to give an average demand rate. The demand for a particular month in the future is then obtained by multiplying this average demand rate by the actual activity for that month, just as with the other techniques.

Techniques 6, 7, 8 and 9 are like 1, 2, 3 and 4 respectively. The only difference is that the inputs are the actual monthly demands

instead of the demands per program element. Hence the D_t and $\hat{d}_{t+k}$ values are the monthly forecasts. The prediction problem is simplified because no estimates of the program element in future months is needed.

Technique 10 is an unweighted nine-month moving average. The procedure is like Technique 5 except that instead of all the data, we use only the most recent 9 months. A 9-month moving average was selected since it gives the same prediction variance as first order exponential smoothing with $\alpha = 0.20$, and also equates the average age of the data.*

For ease of reference, we list the techniques in Table 2.

*See Section III, Eq. 8, or Table 1.

Table 2

PREDICTION TECHNIQUES - DESIGN 1

Technique Number	Description of Technique	Value of Weight α
1	First-order exponentially weighted average applied to demand rates	0.10
2	First-order exponentially weighted average applied to demand rates	0.20
3	Exponential weighting with trend applied to demand rates	0.10
4	Exponential weighting with trend applied to demand rates	0.20
5	Unweighted cumulative average applied to demand rates	----
6	First-order exponentially weighted average applied to actual demands	0.10
7	First-order exponentially weighted average applied to actual demands	0.20
8	Exponential weighting with trend applied to actual demands	0.10
9	Exponential weighting with trend applied to actual demands	0.20
10	Unweighted nine-month moving average applied to demand rates	----

VI. METHODS OF EVALUATING THE TECHNIQUES

THE MEASURES OF ACCURACY

The measures of accuracy used for making comparisons among the techniques are the same as those used in RM-2811, viz., the average monthly error (AME), the relative error (RE), and the root mean square error (RMS). In addition, the average absolute error or mean absolute deviation of the monthly predictions (MAD) was added to this study when we discovered, during discussions of RM-2811, that interest in this measure existed.

To express the above measures in symbolic form, let $\hat{D}_{t+k}$ be the predicted demand in month t+k based on t months of experience. Thus, for example, in Techniques 1 and 2, $\hat{D}_{t+k}$ is D_t multiplied by the activity in the (t+k)th month, whereas in Techniques 6 and 7, it is the same as D_t. Since we are predicting for 12 months in the future, k will take on the values 1, 2, ... 12 for a given t. We let D_{t+k} be the actual demands in the kth month following t months of experience.

With this notation, the four measures of accuracy can be expressed symbolically as follows:

$$AME = \frac{1}{12}\sum_{k=1}^{12} (\hat{D}_{t+k} - D_{t+k})$$

$$MAD = \frac{1}{12}\sum_{k=1}^{12} \left|\hat{D}_{t+k} - D_{t+k}\right|$$

$$RE = \frac{\sum_{k=1}^{12} (\hat{D}_{t+k} - D_{t+k})}{\sum_{k=1}^{12} D_{t+k}}$$

$$\text{RMS} = \sqrt{\frac{1}{12}\sum_{k=1}^{12} (\hat{D}_{t+k} - D_{t+k})^2}$$

There were 21 prediction sets for the B-52 parts and 14 for the Falcon components because of the amount of data available. We further restricted the number of prediction sets by computing the error measures only after "meaningful" experience had occurred. For the B-52 parts we began with predictions based on the first 12 months of experience, and for the Falcon components after 8 months. Hence in the above formulas, t = 12, 13, ..., 21 for the B-52 parts, giving 10 sets of error measures. For the Falcon components, t = 8, 9, ..., 14, yielding 7 sets of error measures.

USING THE MEASURES OF ACCURACY

The four measures of accuracy are summary measures for each line item in the 12-month period following each experience period. Since the Average Monthly Error (AME) is simply the algebraic sum (divided by 12) of the prediction errors for each of 12 months, it can be considered as a measure of the total error for a 12-month period. That is, the computed AME is just one-twelfth of the total error for one year. It seems to be the most appropriate measure for selecting a "preferred" technique because the Air Force is usually interested in predicting requirements over a period of several months, often a procurement leadtime. Of course, any particular measure of accuracy has its disadvantages, but in most cases the technique with the smallest AME also gave the smallest values for the other measures. The preferred technique for each item was selected as in RM-2811. In general,

it was the one which yielded the smallest AME for the greatest number of prediction sets. Modifications of this procedure were made when there was no single technique which satisfied this criterion, as described in the earlier Memorandum.

Our objective, of course, is not to find the best technique for each item. Rather, we want to determine which technique is preferred over a representative sample of items. One obvious approach is to count the number of items for which each technique performs best. However, when many techniques are being compared in this way, it is unlikely that one technique will perform best on a majority of the items. Lacking such a majority, a technique should not be labelled as the "best" simply because it was preferred for the largest number of items. There may be another technique which, when compared with this "best" one, would be selected for a majority of the items. A sufficient test for designating a best technique would be to show that it was preferred for a majority of the items for each paired comparison of techniques. Unfortunately, as the following example shows, there may be no technique with this property. Suppose there are three line items, 1, 2, 3, and three techniques A, B, C. Let technique A be preferred to technique B on two items and let technique B be preferred to technique C on two items. If preferences were transitive, technique A would be preferred to technique C. However, the tabulation below shows that a case can be devised in which C is preferred to A on two parts and therefore no technique is preferred overall. The result can, of course, be generalized to a larger system.*

*This is known as the "paradox of voting" in the literature. See, for example, K. J. Arrow, Social Choice and Individual Values, John Wiley and Sons, Inc., New York, 1951, p. [illegible].

Ranking of Techniques	Line Items		
	1	2	3
Best	A	C	B
Median	B	A	C
Worst	C	B	A

Section IX discusses the defects of this counting procedure at greater length and derives a better method for selecting a preferred technique applicable to the low cost items of Design 2. Since the counting procedure is so simple, however, it was used in both Design 1 and Design 2.

VII. RESULTS OF THE TEST - DESIGN 1

This section summarizes the results of applying the 10 techniques to the two sets of data (B-52 parts and Falcon components) and states our findings and conclusions for Design 1.*

Table 3 shows the number of parts by preferred technique and property class for all 125 parts in the sample, for the 53 parts which had no demands in the first 12 months, and for the 60 parts which had at least 10 demands in the first 21 months. The two sub-samples of 53 and 60 parts respectively were isolated to determine whether some techniques are particularly good on low (high) demand parts.

From Table 3-A we see that Techniques 4, 5, 6 and 9 are preferred for about the same number of parts, as are 1 and 3. There were 50 parts for which exponential smoothing techniques, taking account of flying hours (Nos. 1-4) were preferred, and 50 for which exponential smoothing techniques without program element (Nos. 6-9) were preferred.

We made paired comparisons between Technique 5, the issue rate technique, and Nos. 1, 3, 4, 6 and 9. Many paired comparisons are possible, but they are tedious to make. Since we are interested primarily in comparing the current Air Force procedure with various exponential smoothing procedures which seem to hold some promise of improvement, we limited the number of comparisons. The following numbers of part-preferences were obtained:**

---

*The computer calculations were programmed by D. Hopf.

**The total for each comparison is not 125, the total number of parts in the sample. This is due to the fact that for some parts it was impossible to make a selection based on the two techniques being compared.

BLANK PAGE

NUMBER OF B-52 PARTS BY PH

A. All 1

Property Class	1	2	3	4
Engine	-	-	-	-
Airframe	6	-	4	3
Gunnery	1	1	2	6
Fire Control	-	-	-	1
Communications	5	1	5	3
Accessories	2	-	3	7
Total	14	2	14	20

CHNIQUE AND PROPERTY CLASS

the Sample					
6	7	8	9	10	Total
1	-	-	1	-	3
7	-	1	3	3	33
[illegible]	-	-	12	-	27
[illegible]	-	-	-	-	1
[illegible]	-	2	1	1	31
[illegible]	1	2	5	1	30
[illegible]	1	5	22	5	125

Techniques	Number of Parts
5 1	59 62
5 3	56 68
5 4	66 59

Techniques	Number of Parts
5 6	62 58
5 9	67 56

The difference between the numbers of parts in each comparison is not statistically significant. Hence on the basis of both criteria -- ranking and paired comparisons, we conclude that for this entire sample of 125 parts no one of the exponential smoothing procedures is better than the issue rate procedure currently being used in the Air Force.

Table 3-B shows the preferences when we consider only the 53 parts which had no demands during the first 12 months. That exponential smoothing techniques with trend are preferred to those without trend is logical. Thus Technique 9 is preferred for the largest number of parts, 18. Technique 4 is preferred for the next largest number, 1[illegible]. Both of these use $\alpha = .20$; 9 is applied to actual demands and 4 to demand rates.

Table 3-C shows the distribution of preferences for the 60 parts with at least 10 demands in the first 21 months. Technique 6 has the largest number of preferences, 13; Techniques 4, 5, and 9 each have about the same number. Again we see no real preference for exponential smoothing.

Table 4 shows the number of components of the Falcon missile by preferred technique and general characteristics. Technique 1 was

BLANK PAGE

Tal

Property Class	1	2	3	4
		B.	53 Parts wit	
Engine	-	-	-	-
Airframe	2	-	-	2
Gunnery	1	1	1	3
Fire Control	-	-	-	-
Communications	1	1	2	1
Accessories	-	-	2	5
Total	4	2	5	11
		C.	60 Parts with At	
Engine	-	-	-	-
Airframe	3	-	3	1
Gunnery	-	-	1	3
Fire Control	-	-	-	1
Communications	3	-	2	1
Accessories	2	-	1	3
Total	8	-	7	9

· continued

ique					Total
6	7	8	9	10	
emands in First 12 Months					
-	-	-	-	-	-
2	-	-	1	1	8
-	-	-	11	-	19
-	-	-	-	-	-
-	-	-	1	1	9
-	1	2	5	-	17
3	1	2	18	2	53
10 Demands in First 21 Months					
1	-	-	1	-	3
3	-	1	-	1	12
1	-	-	8	-	14
-	-	-	-	-	1
3	-	1	-	-	15
5	-	-	1	1	15
13	-	2	10	2	60

NUMBER OF FALCON CO
AND GENE

General Characteristics			
	1	2	3
Electronic	2	1	1
Electrical-Mechanical	1	1	2
Mechanical with No Moving Parts	2	-	1
Total	5	2	4

.e 4

; BY PREFERRED TECHNIQUE
ACTERISTICS

que				
5	8	9	10	Total
1	-	-	1	8
3	-	-	-	9
-	1	1	1	6
4	1	1	2	23

preferred for the largest number of components, 5, and Techniques 3, 4, and 5 were each preferred for 4 components. There were no preferences for Techniques 6 and 7. Here we see, however, that to use a program element, check-outs, is better than not to use it: there were 15 preferences for Techniques 1-4, and only 2 for 6-9. This is due to the accelerated phase-in of the Falcon missile during the period in question.*

We made paired comparisons between Technique 5, the issue-rate technique, and Nos. 1, 3, and 4, with the following results.

Techniques	Number of Parts
5	12
1	10
5	12
3	11
5	16
4	7

The difference between the numbers of parts in each comparison is not statistically significant. Hence, as in the case of the B-52 parts, there is no real preference for an exponential smoothing technique, regardless of which criterion is used -- ranking or paired comparison.

*Astrachan, Brown, and Houghton, *op. cit.*, p. 68.

BLANK PAGE

VIII. DESIGN 2 -- DATA, TECHNIQUES, AND RESULTS

The data used in Design 1 were for Hi-Valu and Category II recoverable items for both the B-52 and the Falcon missile. Furthermore, the demands were at base level. It seemed desirable to test exponential smoothing techniques on some Category III items. Demands for such items are generally much higher than for Hi-Valu and Category II-R parts; there are many more Category III parts; and they are managed differently. If applicable, the automatic computing procedures of exponential smoothing (with or without trend) and decreased data storage requirements could be valuable.

Oklahoma City Air Materiel Area (OCAMA) had issue history on about 160.000 WSM* B-52 items for the 29½ months from April 16, 1960 to September 30, 1962. Data on a sample of about 16,000 items were made available to us. The information for each line item included, among other things, the number of requisitions and the number of issues on a quarterly basis, unit cost, and the Expendability/Repair/Cost (ERC) Code. Expendable Category III items have ERC code "N".

We selected a sample from the 16,000 consisting of every 25th item that had some issues during the first seven quarters and ERC code N, beginning with the first item in the listing that satisfied these conditions. There were 425 items in this sample. On the grounds that exponential smoothing would probably not be used on items with less than one demand per quarter, we then eliminated all items which had fewer than 10 issues during the first seven quarters. This reduced our sample size to 292.

---

*Weapon System Manager -- under this management concept most of the issues are made to satisfy routine base demand. For items which are peculiar to the weapon system there will be some issues to Specialized Repair Activities (SRA).

It is of some interest to examine the fluctuations in the average issues per line item during the data-collection period, as shown in Table 5.

The average number of issues per line item for all ten quarters is 163.4. For the first 4 quarters it is 148.5; for the first 7 it is 162.0. The largest average occurs in the seventh quarter, 239.1, and the smallest in the first, 104.4.* There is no pattern to these averages Thus the average number of issues per month ranged from

Table 5

AVERAGE ISSUES PER LINE ITEM
(Ten quarters, 292 items)

Quarter	Average
1	104.4
2	143.9
3	185.5
4	160.2
5	137.7
6	162.9
7	239.1
8	154.1
9	145.9
10	200.3

about 3 to 8 per line item. In our sample of 292, there occurred 51 items for which there were recorded issues during only one of the first seven quarters. Some of these had issues during the last 3 quarters. For all 425 items in the original sample, the averages are reduced by about one-third.

---

*This may be due in part to the fact that the data for what we are calling the first quarter really were for only 2½ months. We feel that the effect of this on our subsequent work is negligible.

The most expensive item in our initial sample of 425 cost $159. There were 4 issues of it during the fourth quarter and none at any other time. The next most expensive unit cost $139. Thirty-five of these were issued during the first 7 quarters and 30 in the last 3. There was one issue in the fourth quarter and none at any other time for an item which cost $99. Twenty $98.50 items were issued during the 5th quarter and 18 in the last 3. All the other items cost less than $90. Only the second and fourth of these four items were included in our sample of 292.

There were 14 of the cheapest item, which cost one cent. Issues ranged from 6 to 53,698 during the first 7 quarters and from none to 18,742 in the last 3 quarters. There were 5 of these items which had no issues during the last 3 quarters. During the first 7, they had 402, 450, 555, 599, and 4,005 issues. The item which had six issues in the first 7 quarters had 102 in the last three. It was not included in our sample of 292.

The above again emphasizes the difficulties inherent in demand prediction due to the irregularities in demand patterns. This last bears out the results of earlier RAND studies.*

In order to give the reader some feeling for the data, we have included in Table 6 the unit cost and quarterly issues for every 10th item in our sample of 425. It can be seen that issues from depot are erratic. In fact, the ratio of variance to mean is greater than one for nearly all the parts in the table. For some parts it is greater than 50, as for example, numbers 22 and 31.

---

*See, for example, Bernice Brown, Characteristics of Demand for Aircraft Spare Parts, The RAND Corporation, R-292, July, 1956.

Table 6

UNIT COST AND QUARTERLY ISSUES FOR A SAMPLE OF CATEGORY III B-52 PARTS

Part No.	Unit Cost ($)	Quarters									
		1	2	3	4	5	6	7	8	9	10
1	4.50	0	0	0	0	0	0	10	0	0	0
2	.50	0	2	1	2	0	0	0	0	0	0
3	17.00	0	0	0	0	0	1	0	1	2	0
4	4.60	0	0	1	0	0	0	0	0	0	0
5	62.00	0	0	4	0	0	0	0	0	0	0
6	.99	0	0	0	5	0	0	0	0	0	0
7	7.25	0	0	0	0	0	0	11	0	0	0
8	4.00	0	49	16	6	0	10	4	6	38	35
9	9.45	0	4	0	12	0	0	1	0	2	0
10	6.25	1	0	0	1	0	0	0	1	2	0
11	5.95	0	0	0	10	10	3	30	15	14	65
12	16.25	0	0	0	2	2	3	0	0	0	0
13	.30	0	0	0	0	9	33	0	0	5	0
14	4.65	0	0	21	1	15	4	7	4	1	5
15	.06	0	0	0	100	100	500	0	0	400	0
16	.75	0	0	1	0	2	0	6	4	0	2
17	11.00	0	0	1	1	0	0	0	0	0	0
18	.03	0	0	0	21	0	21	50	0	0	0
19	.02	0	0	550	350	200	350	0	0	0	150
20	.27	224	26	6	223	96	0	0	0	20	0
21	2.60	0	0	10	30	0	0	10	12	6	6
22	.13	0	0	0	0	0	60	0	0	0	0
23	.01	0	0	6800	7867	3400	11301	24330	2151	5235	11356
24	.01	0	104	121	98	0	0	0	0	8	0
25	2.85	10	9	1	39	60	2	55	0	31	9
26	.90	0	7	0	50	0	0	0	0	0	0
27	.11	764	1093	302	232	559	1109	380	659	1535	122
28	.80	0	0	0	0	0	0	61	7	19	300
29	1.80	0	0	0	30	0	0	0	0	0	0
30	.03	0	0	0	0	3	2	0	0	2	0
31	.40	0	80	0	0	0	0	0	0	0	3
32	.14	0	0	0	0	0	10	16	1	5	0
33	67.00	0	0	0	0	0	0	10	0	0	1
34	1.00	0	0	0	0	0	16	0	0	0	0
35	1.65	0	0	0	0	0	0	11	0	0	0
36	3.00	0	0	0	0	0	1	0	0	0	0
37	4.00	0	1	4	2	4	5	2	1	26	5
38	4.25	0	0	0	0	0	0	1	0	0	0
39	14.00	0	0	0	4	5	0	2	2	0	0
40	1.05	0	6	0	0	4	0	0	15	0	1
41	3.00	34	0	26	121	16	102	102	0	9	1
42	5.00	0	0	0	0	0	0	2	0	0	0

There were 18 property classes represented in our sample of 292, only four of which had at least 10 line items in them. We consolidated the remaining 14 classes in stating our results. The following table gives the number of items included in the property classes in our sample:

FEDERAL SUPPLY GROUP (FSG)	No. of L/I in Sample
15 -- Aircraft and airframe structural components	42
47 -- Pipe, tubing, hose, and fittings	22
53 -- Hardware and abrasives	115
59 -- Electrical and electronic equipment components	71
All others	42
Total	292

The test design for these data differed from the one employed for the B-52 and Falcon data. We used six techniques and predicted for three quarters in the future. Historical base periods consisting of four quarters of data and seven quarters of data were used.* No program element was introduced. The techniques are defined as follows:

Technique Number	Base Period	Technique
1	first 4 quarters	Issue rate
2	first 4 quarters	Exponential smoothing, $\alpha = .20$
3	first 4 quarters	Exponential smoothing, $\alpha = .30$
4	first 7 quarters	Issue rate
5	first 7 quarters	Exponential smoothing, $\alpha = .20$
6	first 7 quarters	Exponential smoothing, $\alpha = .30$

The same four measures of accuracy were used as in the first part of this study. Selections were made independently for the four-quarter base and the seven-quarter base. The number of line items for which each technique is preferred is given in Table 7.

*The initial estimate for the exponential smoothing technique is obtained by averaging the data over the base period. (See Eq. 3.)

Table 7

NUMBER OF PARTS BY PREFERRED TECHNIQUE AND FEDERAL SUPPLY GROUP

FSG	Technique: Four Quarter Base 1	2	3	No Preference	Seven Quarter Base 4	5	6	No Preference	Number of Line Items
15	7	--	6	29	12	--	14	16	42
47	5	1	5	11	5	1	10	6	22
53	34	2	48	31	50	6	43	16	115
59	10	1	20	40	25	2	21	23	71
Others	6	--	9	27	15	1	13	13	42
Total	62	4	88	138	107	10	101	74	292

Using four quarters of historical data, we were unable to identify a preferred technique for 138 parts (47 per cent). Even when we used 7 quarters of data, 74 of the 292 parts (25 per cent) indicated no preferred technique.

Technique 3 ($\alpha = .3$) is preferred for the largest number of parts based on 4 quarters of data. Yet there is no outstanding preference when 7 quarters are used, Techniques 4 (issue rate) and 6 ($\alpha = .3$) being preferred by about the same number.

It is interesting to examine the technique preferences for a line item based on 4 and 7 quarters of data. Hopefully, if Technique 1 were preferred for a part based on the first 4 quarters of data, then Technique 4 should be preferred based on the first 7 quarters of history. The same should hold true for Techniques 2 and 5, and 3 and 6. This did not happen, as can be seen from Table 8.

Table 8

NUMBER OF PARTS BY PREFERRED TECHNIQUE BASED ON 4 AND 7 QUARTERS OF DATA

Technique - 4 Quarters	Technique - 7 Quarters				
	4	5	6	No Preference	Total
1	12	3	32	15	62
2	2	--	1	1	4
3	42	5	35	6	88
No Preference	51	2	33	52	138
Total	107	10	101	74	292

There were only 12 parts for which the issue rate technique was preferred regardless of the amount of data, none for 2 and 5, and 35 for 3 and 6 (exponential smoothing, $\alpha = .3$). There were 52 parts (18 per cent) for which there were no preferences based on either 4 or 7 quarters of data. Thus we see that if one technique is preferred based on 4 quarters of data, there is no reason to suppose that this same technique would be preferred as additional demand experience becomes available.

Techniques 1, 2, and 3 tended to overestimate as a group and Techniques 4, 5, and 6 tended to underestimate as a group. The

following table shows the number of parts for which the preferred technique under- and overestimated. Where there was no preference, it was still possible to categorize the parts in this respect.

Table 9

NUMBER OF PARTS UNDERESTIMATED OR OVERESTIMATED BY PREFERRED TECHNIQUE

4 Quarters	7 Quarters: Underestimate	Overestimate	Total
Under-estimate	72	110	182
Over-estimate	34	76	110
Total	106	186	292

The number of parts for which the techniques, using the two bases, consistently underestimated or overestimated was about the same, 72 and 76, respectively. For the totals, the numbers of line items were reversed, 182 and 110, compared with 106 and 186.

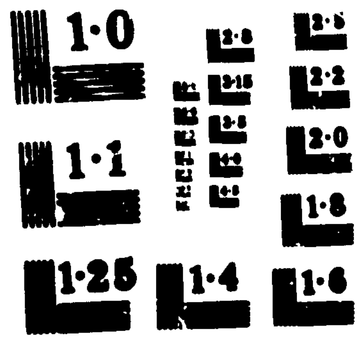
We did not use second order smoothing because it is unlikely that one would want to use a prediction technique with a trend factor of only 4 quarters of data, or even of 7 quarters of data. Furthermore, there is usually some question about the accuracy of the information at the outset of a data collection program.

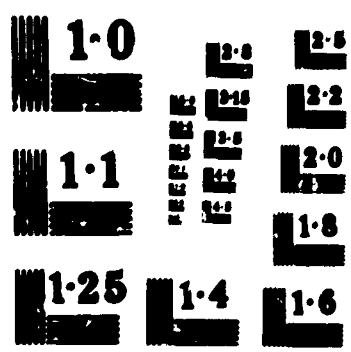
We also examined the 133 parts in our original sample of 425 which had fewer than 10 issues during the first 7 quarters. With a base period

of 4 quarters of data, there was no preferred technique for 124 of them. Of the remainder, Technique 1 was preferred for 8 of the items and Technique 3 for 1 of them. When we used 7 quarters of data to predict, 113 items exhibited no preference, Technique 4 was preferred for 6 and Technique 6 for 14 items.

Our data included the number of requisitions by quarter, as well as the number of issues. We applied our techniques to predict the number of requisitions per quarter. Using 4 quarters of data, we predicted the number of requisitions for the next 3 quarters and multiplied these figures by the number of issues per requisition during the base period to get the number of demands. Then we followed the same procedure using 7 quarters as a base. The results of the selection of preferred techniques were substantially the same as when we did not use requisitions. Hence the results are not included here.

CONT.





BLANK PAGE

IX. MEASUREMENT OF AGGREGATE LOSS USING A STATISTICAL DECISION THEORETIC APPROACH

In Sec. VI we described four measures of error that were used to assess prediction accuracy: average monthly error, average absolute error, root mean square error, and relative error. Using the B-52 and Falcon data, we observed that for a particular line item the prediction technique that resulted in the smallest value of average monthly error usually had the smallest value for the other three measures of error as well. We selected as the preferred one for each item that technique with the smallest average monthly error for the greatest number of prediction sets. To determine the preferred technique over the entire sample of 125 B-52 parts and 23 Falcon components we simply counted the number of times each technique was preferred (see Tables 3 and 4). Similar procedures were applied to the OCAMA data.

Such a counting procedure has several drawbacks. In the first place, each item is weighted equally though the average demand and unit cost vary widely from item to item. Secondly, if no technique performs best on a majority of the items, one should evaluate them by pairs. If there are several techniques, the number of required comparisons may be very large. Still worse is the possibility that there is no preferred technique, since the preference relations are not transitive.* Thirdly, the procedure is insensitive to the magnitude by which a technique is preferred. For example, the counting procedure will label as "preferred" a technique that performs slightly better than a second technique on the majority of items but far worse on the balance.

---

*See Sec. VI.

We shall now describe another measure of aggregate error suitable for use on the OCAMA data.* This measure does not possess the three defects of the counting procedure noted above. Since we are concerned with low-cost items that are managed under economic order quantity procedures, the approximate expression** for total variable cost (TVC) for a line item during a time period is:

$$\text{(13)} \qquad TVC = \frac{Q}{2} IC + \frac{x}{Q} S ,$$

where

Q = economic order quantity

I = interest rate per period

C = unit cost

x = demand per period

S = cost of placing an order

This is the familiar cost expression underlying the classical Wilson economic lot size formula.*** The first term on the right-hand side is the holding cost per period and the second term is the procurement cost per period. By differentiating with respect to Q and setting the derivative equal to zero, we obtain the well-known result for minimum cost,

$$\text{(14)} \qquad Q_{optimum} = \sqrt{\frac{2xS}{IC}} .$$

In Eqs. 13 and 14 we have assumed that x, the true value of demand during the period, is known. Actually our problem is to make an

---

*This procedure was suggested by G. J. Feeney.

**We are assuming x and Q are continuous variables.

***See for example T. M. Whitin, The Theory of Inventory Management, Princeton University Press, 1957, or A. R. Ferguson and Lawrence Fisher, Stockage Policies for Medium and Low-Cost Parts, The RAND Corporation, RM-1962, April 1958.

estimate of the value of demand which we shall denote by $\hat{x}$. This estimated value $\hat{x}$ will be used to determine Q in Eq. 14. This value of Q is then substituted into Eq. 13 to give us the cost. The number of procurements, x/Q, and hence the cost, will still depend on the true value of demand, x.

Let us define $u(\hat{x},x)$ as the cost during a period when demand is estimated as $\hat{x}$ but is actually x. Then from Eqs. 13 and 14,

$$(15) \qquad u(\hat{x},x) = \sqrt{\frac{ICS}{2}}\left\{\sqrt{\hat{x}} + \frac{x}{\sqrt{\hat{x}}}\right\}.$$

It is easily shown that this is minimized when the random variable $\hat{x}$ assumes the value x.

Let us define a loss function, $L(\hat{x},x)$ as the cost when our estimate of demand is $\hat{x}$, minus the cost if we had made the correct decision $\hat{x} = x$. Then

$L(\hat{x},x)$ = loss during a period when demand is estimated as $\hat{x}$ instead of x.

$$= u(\hat{x},x) - u(x,x)$$

$$(16) \qquad = \sqrt{\frac{ICS}{2}}\left\{\sqrt{\hat{x}} + \frac{x}{\sqrt{\hat{x}}} - 2\sqrt{x}\right\}.$$

Of course, the true value x is not known, but we shall assume that it is distributed according to a probability distribution q(x). Our problem is to choose $\hat{x}$ so that the expected value of the loss function is minimized, i.e., so that

$$(17) \qquad E\,\{L(\hat{x},x)\} = \int_x L(\hat{x},x)q(x)dx$$

is minimized.

Substituting Eq. 16 into Eq. 17 and taking the partial derivative with respect to $\hat{x}$ yields the minimizing condition

$$\sqrt{\frac{ICS}{2}} \int_x \left[\frac{1}{2}\, \hat{x}^{-1/2} \; -\frac{1}{2}\, \hat{x}^{-3/2} \cdot x\right] q(x)dx \quad = 0,$$

whence

$$\hat{x} = \int_x xq(x)dx \; . \tag{18}$$

In other words the loss function will be minimized if we choose an estimate, $\hat{x}$, equal to the mean of the distribution on the true parameter. Thus it is not necessary to know anything about the distribution $q(x)$ except its mean. Hence, the exponential smoothing and moving average techniques which estimate the mean are consistent with this loss function.

In Fig. 2 we have plotted the loss function of Eq. 16 divided by $\sqrt{x}$ against the ratio of estimated to observed demand, $\frac{\hat{x}}{x}$, assuming $\sqrt{\frac{ICS}{2}} = 1$. Note that the graph is not symmetric. If we had plotted the loss function against log $\frac{\hat{x}}{x}$, the graph would be symmetric about $\frac{\hat{x}}{x} = 1$.

By considering x to be a constant we can see the effect of changes in the ratio $\hat{x}/_x$. That the value of $L(\hat{x},x) / \sqrt{x}$ is greater for an underestimate than for an overestimate of the same amount agrees with our intuition of how the function should behave in an inventory system. If we over- or underestimate by the same relative amounts, the values of the function are the same.

The aggregate measure of loss for a technique is obtained by simply computing Eq. 16 for each item and summing over all items. Note that the

dimensions of Eq. 16 are dollars, and that consequently this aggregate loss has an economic interpretation which is more useful than a measurement of aggregate error in per cent, for example. The value of $\hat{x}$ is obtained from the prediction technique, and the value of observed demand

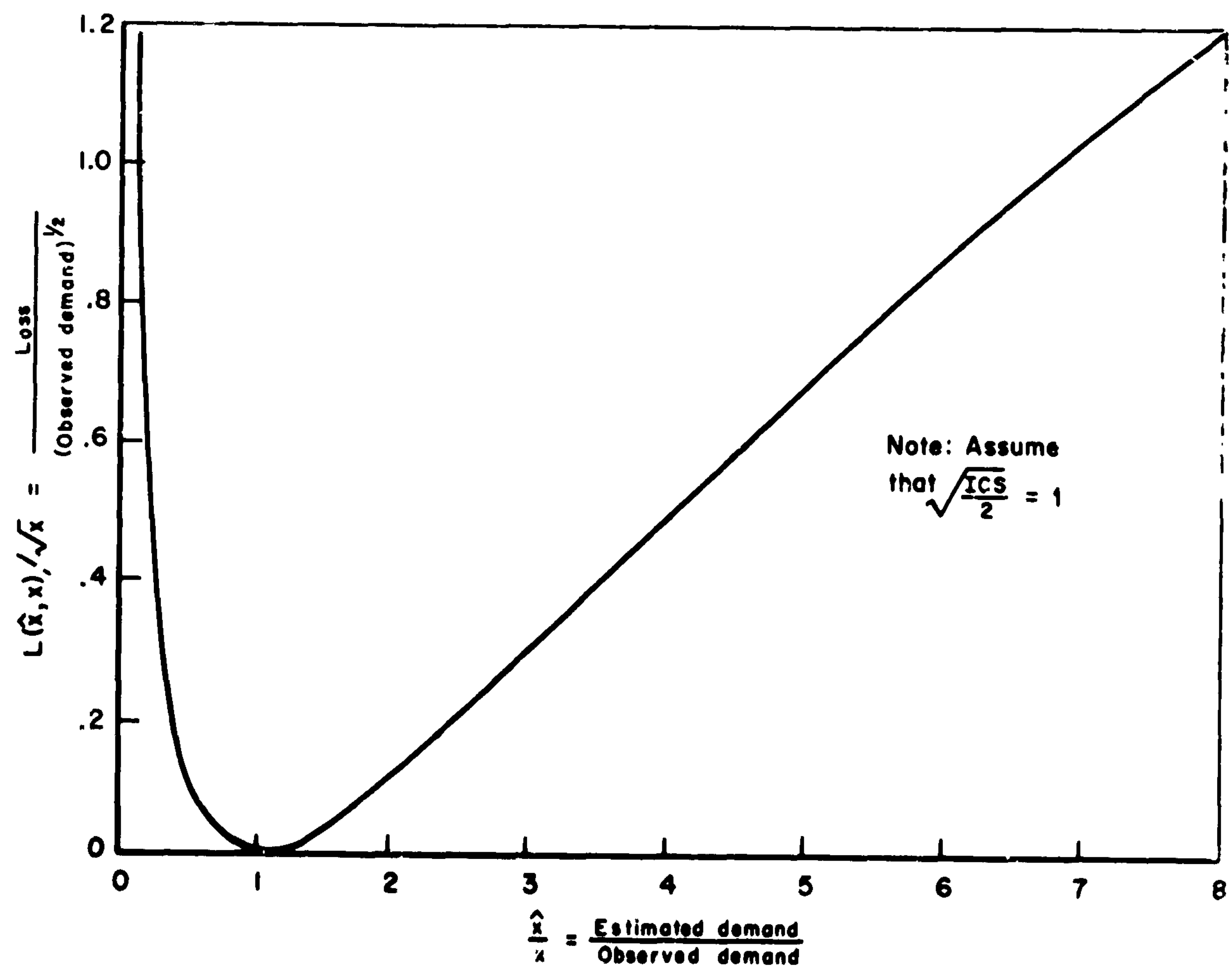


Fig. 2 — The loss function divided by $\sqrt{x}$

is entered for x. We artificially set $\hat{x} = 1$ in those cases where the prediction technique would have specified zero to prevent Eq. 16 from becoming

infinite. Over a group of items, I and S will usually be considered constants. Since the statistical analysis which follows is not affected by these values* we assumed that $\sqrt{\frac{IS}{2}} = 1$.

For the entire sample of 425 line items, the aggregate loss/item during the prediction period under each technique is given in the following tables:

Four Quarter Base	
Technique	Aggregate Loss/Item
1	$ 9.22
2	9.18
3	9.05

Seven Quarter Base	
Technique	Aggregate Loss/Item
4	$ 4.88
5	4.58
6	4.46

We can also think of the entries in the second column in each table as estimates of the average dollar loss/item (assuming $\sqrt{\frac{IS}{2}} = 1$) per leadtime period for increased costs of procurement and holding due to incorrect demand forecasts.**

We cannot compare Techniques 1, 2, and 3 as a group with Techniques 4, 5, and 6, because of the difference in the base and prediction periods.

---

*One reasonable set of values satisfying $\sqrt{\frac{IS}{2}} = 1$ is to choose I = .10 per period and S = $20 per order. In this case the period is 3 quarters so that the yearly interest rate would be (4/3) .10 = .13. Studies performed by M.I.T. for the Army Ordnance Corps indicate that yearly interest rates of .17 and depot order costs of $100 are reasonable in that application. Under these assumptions, the value of $\sqrt{\frac{IS}{2}} = 2.53$.

We used the sign test, t-test, and Wilcoxon test to determine whether the results were statistically significant. These tests are not affected when the variable is multiplied by a constant, in this case $\sqrt{\frac{IS}{2}}$.

**It is perhaps of some interest to note that during quarters 5, 6, and 7 for which the first three techniques predict, the dollar value of issues/line item was $161.43; during quarters 8, 9, and 10 for which the last three techniques predict, the dollar value of issues/line item was $144.83.

It is admissible, however, to compare these two groups of techniques within themselves. The application of statistical tests indicated that there were no significant differences among 1, 2, and 3 or among 4, 5, and 6 using the entire sample of 425 items.

We also applied the above procedures to the reduced sample of 292 items with at least ten demands in the first seven quarters. The results were the same.

As a final test we predicted the demand for the last three quarters based on the preceding four quarters of data. In other words we used Techniques 1, 2, and 3, but with quarters 4-7 instead of 1-4 as a base. This enabled us to compare the aggregate loss based on four quarters with the aggregate loss based on seven quarters (using Techniques 4, 5, and 6), since the period being predicted was the same. The aggregate loss/item was found to be \$4.96, \$4.86, and \$4.79 to correspond with techniques 4, 5, 6, respectively, in the table above. We note that the use of three additional quarters of data on which to base predictions decreases the aggregate loss/item. These decreases are not statistically significant. However, we are led to the speculation that data which are more than a couple of years old may have negligible value for prediction purposes. More precisely, it appears likely that the information, if any, provided by data older than, say, two years is related primarily to the non-stationary characteristics of the demand distribution.

In conclusion, we should remind the reader that the loss function discussed in this section considers only procurement and holding costs. We realize that stockout costs, which we have omitted, are extremely important, but they are difficult to assess objectively. How much does

a stockout cost? Is a stockout on a $ .01 item as costly as a stockout on a $30.00 item? Are ten stockouts on an item ten times as costly as one stockout?

The advantage of avoiding arbitrary assumptions about stockout costs is that the loss function in Eq. 16 has only two parameters, I and S, which are constant over all the items. We can make reasonable estimates of these parameters. Further, since they are constant for all items, they do not affect tests of statistical significance based on the loss functions. Finally, it is important to keep in mind that the aggregate loss function is being used to evaluate prediction techniques. We are not constructing an inventory policy. If we were, we should be obliged to consider stockouts.

X. FINDINGS AND CONCLUSIONS

Exponential smoothing does not appear to be a significantly better predictor than the cumulative issue rate techniques currently being used. The ranking procedure used to evaluate both designs did not enable us to pick a preferred technique. When we applied the more sensitive test of aggregate loss in Design 2 for the Category III data, there were still no statistically significant differences between techniques.

We remind the reader, however, that exponential smoothing has definite computational advantages. For first order exponential smoothing only one average need be stored for each item -- this in contrast with the requirements of a cumulative issue rate or a moving average. A trend can be readily accommodated with exponential smoothing, and the rate of response due to the smoothing constant can be easily changed.

A measurement of aggregate loss seems to be of fundamental importance in assessing prediction accuracy. Some readers may complain about a sophisticated measurement of error such as the aggregate loss function developed in this Memorandum using a statistical decision theory point of view. Obviously it is based on a simple model which balances procurement costs against holding costs. It is certainly not a comprehensive model. On the other hand it is clear that simple ranking procedures are highly questionable and furthermore that they are not sufficiently sensitive. Ranking procedures ignore magnitude information, behaving like a sign test in classical statistics. We feel that future empirical tests of prediction techniques will inevitably fail unless sensitive measuring instruments such as aggregate loss functions are employed. Such functions are less arbitrary and logically more appropriate.

Program element data are valuable for prediction. This conclusion is based on the analysis of the Falcon data in Design 1. Prediction accuracy was substantially improved for 19 of 23 parts by the inclusion of a program element (checkouts). Of course, the use of a program element requires that it be forecasted so that predicted demand per program element can be converted to predicted demand. Naturally the importance of a program element is determined by the rate at which a weapon is phased in.

No program element information was available for the depot issues of Category III items in Design 2.

The use of requisition data did not alter the accuracy of demand prediction. For the Category III depot issues of Design 2 we had quarterly data by item giving total requisitions and total issues. When the techniques of Design 2 were applied to requisitions for prediction and then multiplied by the average issues per requisition computed over the base period, the resulting predictions were substantially unchanged. Of course, there is an unlimited number of ways that the requisition data could have been used. For instance, we might have applied the techniques to issues per requisition, but this would require an estimate of requisitions in the future by item. Such a procedure would be similar to using a program element except that it would necessitate a different program element for each item. Since this does not seem feasible, we restricted our attention to the one application of the requisition data described above. Our conclusion that requisition data did not alter the accuracy of prediction is based on this application. However, we do not exclude the possibility that a

reasonable procedure may be devised that will be able to extract information from the requisition data for prediction purposes.

The accuracy of first order prediction does not increase substantially when the base period becomes longer than a year. This conclusion is based on the depot issues for Category III items in which the aggregate loss/item was decreased by a maximum of 7 per cent (on exponential smoothing with a constant of .30) when the base period was extended from four quarters to seven. Trend calculations (second-order) were not made because of the paucity of data. Our conclusion suggests that the information, if any, in data more than a couple of years old is primarily related to trend.

The variance in depot issues is extremely high. If demands for an item from the depot were placed in a random manner, statistical theory predicts that we would observe a variance to mean ratio of one in the demand pattern. The variance to mean ratios that we observed in the Category III depot data were almost always greater than one, and often as large as 50 or 100. Actually we know that demands on the depot do not occur at random by design. Bases order large quantities at infrequent intervals according to an economic lot-size type of criterion that attempts to balance the costs of procurement with the costs of holding. This artificially amplifies the fluctuations of demand that are made on the depot. Very little empirical work has been done for multi-echelon demand problems, but it is obvious that an inventory system designed to optimize base performance only may be decidedly non-optimum on a system basis.

BIBLIOGRAPHY

Arrow, K. J., Social Choice and Individual Values, John Wiley and Sons, Inc., New York, 1951.

Astrachan, Max, Bernice Brown, and J. W. Houghten, A Comparative Study of Prediction Techniques, The RAND Corporation, RM-2811, December, 1961.

Brown, Bernice, Characteristics of Demand for Aircraft Spare Parts, The RAND Corporation, R-292, July, 1956.

Brown, R. G., Smoothing, Forecasting, and Prediction of Discrete Time Series, Prentice-Hall, Englewood Cliffs, N.J., 1963.

Brown, R. G., and R. F. Meyer, "Fundamental Theorem of Exponential Smoothing," Journal of the Operations Research Society of America, Vol. 9, No. 5, Sept.-Oct., 1961, pp. 673-687.

Campbell, H. S., The Relationship of Resource Demand to Airbase Operations, The RAND Corporation, RM-3428-PR, January, 1963.

Dobbie, J. M., "A Simple Proof of a Theorem in Exponential Smoothing," Journal of the Operations Research Society of America, Vol. 11, No. 3, May-June, 1963, pp. 461-463.

Ferguson, A. R., and Lawrence Fisher, Stockage Policies for Medium and Low-Cost Parts, The RAND Corporation, RM-1962, April, 1958.

McGlothlin, W. H., Development of Bayesian Parameters for Spare Parts Demand Prediction, The RAND Corporation, RM-3699, July, 1963.

Whitin, T. M., The Theory of Inventory Management, Princeton University Press, Princeton, N.J., 1957.